



X Coloquio Internacional sobre Tecnologías Aplicadas a los Servicios de Información

Universidad Nacional Experimental de Táchira
Venezuela. Noviembre 2010

Servicios de Información Digital Sociales y Semánticos: Nuevos paradigmas 3.0

Eva M^a Méndez Rodríguez

emendez@bib.uc3m.es

Dpto. Biblioteconomía y
Documentación



Universidad Carlos III
de Madrid

San Cristóbal (Venezuela), 17 de Noviembre de 2010





Fuente: <http://media.photobucket.com/image/UNET/BEKERMANN7/unet.jpg>



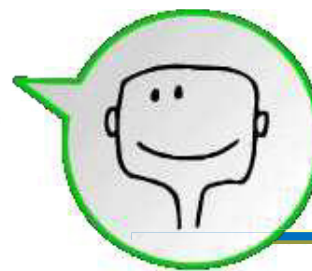
De qué querría hablar...



... de muchas cosas

- De nuestro mundo de los contenidos digitales
- De “la gestión integral de unidades de información”
- De cómo crear, gestionar, recuperar, etc. objetos de información digital
- De modelos de datos, estándares, etc. para organizar sistemas de información digital
- De vocabularios y la nueva semántica de la WWW (ontologías... folksonomías)
- De paradigmas y denominaciones: 2.0, 3.0, *paradigma 2020*.

De qué vamos a hablar



De todo esto... pero centrándonos en las nuevas tendencias de sistemas y servicios de información digital:

- El problema de la representación y recuperación de información en la WWW (del *data mining* a las ontologías)
- De la Web Semántica
- Web 2.0 y la Web Semántica: Web 3.0... Web Social Semántica (*Social Semantic Web*): Folksonomías + Ontologías
- *Linked Open Data*... y de “la nube”
- Nuevos perfiles profesionales y los Servicios de Información Digital Sociales y Semánticos.



I know
what
you
mean...

THE SEMANTIC WEB

A new form of Web content
that is meaningful to computers
will unleash a revolution of new possibilities

by
TIM BERNERS-LEE,
JAMES HENDLER and
ORA LASSILA

PHOTOILLUSTRATIONS BY MIGUEL SALMERON

WILEY 2000



Miles de objetos digitales ...

LA WEB (Caracterización)

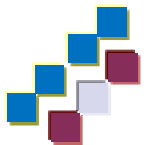


- WEB: Gran cantidad de información (de alta calidad e información científica y/o profesional + información informal, sitios personales, Weblogs, Wikis, etc.)
- ¿Cómo dotar a esa información de **fiabilidad** y consistencia en la recuperación?
- Además de la información nacida en la Web (*born digital*) → **tendencia al acceso, disponibilidad en línea del patrimonio cultural / científico** (EU, España, WDL, etc.)

Lo que no está en la Web... no existe

Lo que no se encuentra, se agrega o se integra...

... no vale



Miles de objetos digitales / SSID



Digitalización masiva de documentos



Producción distribuida de contenido digital (born digital)



Archivos, bibliotecas... Servicios de Inf. Digital



<http://pares.mcu.es>

<http://hispana.mcu.es>



Museos... Servicios de Inf. Digital



ceres
Colecciones en Red

Inicio | Ayuda

Búsqueda General

Búsqueda Avanzada

Selección por agrupaciones de Museos

Selección de Museos

Inventario

Clasificación Genérica

Objeto/Documento

Título

Autor

Emisor

Materia/Soporte

Iconografía

Contexto Cultural/Estilo

Datación

Lugar de Producción/Ceca

Lugar de Procedencia

Lugar Específico/Yacimiento

<input type="text"/>
<input type="text"/>
<input type="text"/>
<input type="text"/>
<input type="text"/>
<input type="text"/>
<input type="text"/>
<input type="text"/>
<input type="text"/>
<input type="text"/>
<input type="text"/>
<input type="text"/>
<input type="text"/>

Seleccionar Término

Seleccionar Término

Seleccionar Término

Seleccionar Término

Seleccionar Término

Seleccionar Término

Seleccionar Término

Seleccionar Término

C. Fechas/Periodos

Seleccionar Término

Seleccionar Término

Seleccionar Término

Jerárquica

Jerárquica

Jerárquica

Jerárquica

Jerárquica

Jerárquica

Jerárquica

C. Avanzada

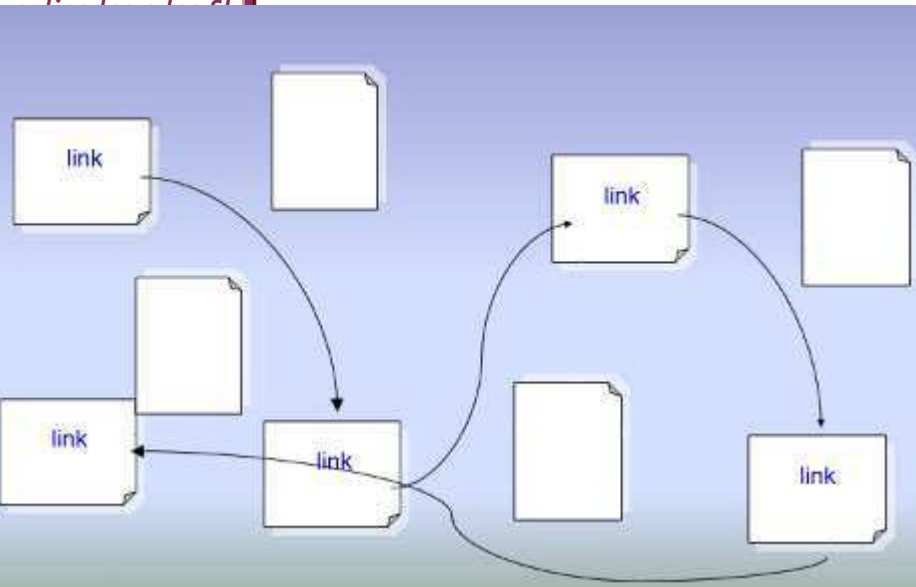
Jerárquica

Jerárquica

Mostrar resultados en mosaico de imágenes

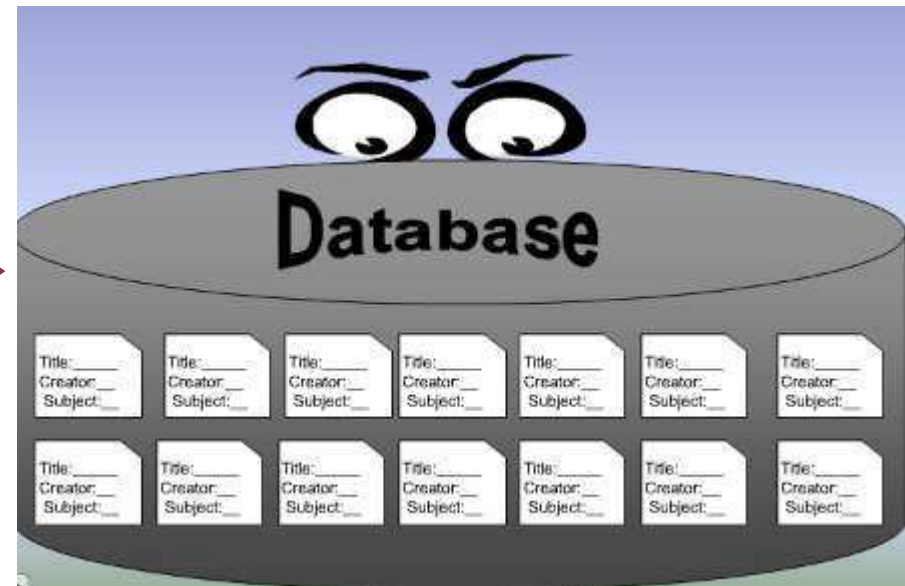
Buscar Limpiar

Web de Documentos / Web de Datos



Web de documentos

Web de datos

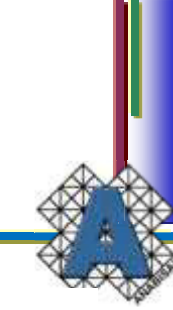


Miles de objetos digitales: Nuevos retos



- **La obsolescencia tecnológica y la necesidad de preservar/custodiar el acceso al patrimonio digital. Tendencias:**
 - *Digital curation* (Gestión de documentos... ciclo de vida-DCC)
 - *Digital preservation* (Bibliotecas digitales, digitalización, long-term preservation)
 - *Web archiving*
- **Representación y recuperación de información digital. Tendencia:**
 - **Web Semántica y LOD**

La Web Semántica... más...



Lo que Internet necesita es un/a bibliotecario/a anticuado/a.

Encontrar lo que queremos en la Web debería de ser igual de fácil que encontrar un libro en una biblioteca.

Así será, si Eugenia [pongan aquí su nombre] consigue lo que quiere. Está trabajando para crear un **estándar** para etiquetar y catalogar información online - que incluye todos los 2.7 billones de páginas web... algo así como un sistema de información Decimal Dewey virtual, si lo preferís. Así, podrás gastar **menos tiempo en buscar, y más tiempo en utilizar la información** que necesitas.

Shhhhh... Estás en Internet.





Representación y Recuperación de la información Digital: el problema de siempre y nuevas tendencias



Coloquio Internacional
sobre Tecnologías aplicadas
a los Servicios de Información
"Hacia la Gestión Integral de Unidades de Información"

17-18 y 19/Noviembre, 2010

El problema de la representación y recuperación de información



- Problema común: organización, gestión y recuperación de información:
 - Publicación de contenidos
 - Acceso a la información y cambio (mutabilidad) de la información
 - Integración de recursos distribuidos
 - Categorización y valoración de contenidos
 - Confianza o fiabilidad de los contenidos.
- Solución tradicional (Google, etc.): aunque han mejorado mucho... no suficientes para recuperar **conocimiento**



El problema de la representación y recuperación de información



- Ambigüedad en el significado de las cadenas de búsqueda
- Multitud de resultados sin orden lógico (Google)
- Imposible trasladar la relevancia subjetiva del usuario (incapacidad de entender el significado que el usuario tiene en mente)
- No garantía de fiabilidad
- Falta ayuda para la formulación de las búsquedas



Representación/Recuperación de Información: Data mining



- Recuperar conocimiento no es lo mismo que recuperar información
- Recuperar conocimiento implica tradicionalmente el procesamiento semántico de grandes cantidades de datos: Data mining
- El *data mining* (minería de datos, explotación de datos, etc.) es un tema complejo (informática + técnicas computacionales estadísticas, reconocimiento de patrones...



Representación/Recuperación de Información: Data mining



- *Data mining* → extracción de conocimiento de grandes cantidades de información estructurada
- *Text mining* → extracción de información de calidad a partir de textos o información no estructurada.

“Buscar” vs. “Recuperar (descubrir)”



Búsqueda
(objetivo)

Descubrimiento
(oportunista)

Datos
Estructurados

Data
Retrieval

Data
Mining

Datos sin
Estructurar (texto)

Information
Retrieval

Text
Mining





La Evolución de la Web

... Siempre estamos hablando de la Web del Futuro



Coloquio Internacional
sobre Tecnologías aplicadas
a los Servicios de Información
"Hacia la Gestión Integral de Unidades de Información"
17-18 y 19/Noviembre, 2010

La Web en el FUTURO (2001)



- Continuará la mezcla de calidades informativas:
 - Se mantendrán sitios personales
 - Escaparates comerciales con un propósito real
- Los ordenadores utilizarán información estructurada de forma mucho más eficaz.
- Se necesita la integración de datos para permitir el procesamiento automático
- SW (Web Semántica)...

“...is an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation”

Tim Berners-Lee, James Hendler, Ora Lassila. The Semantic Web. Scientific American (Mayo 2001)



La Web en el FUTURO (2005)



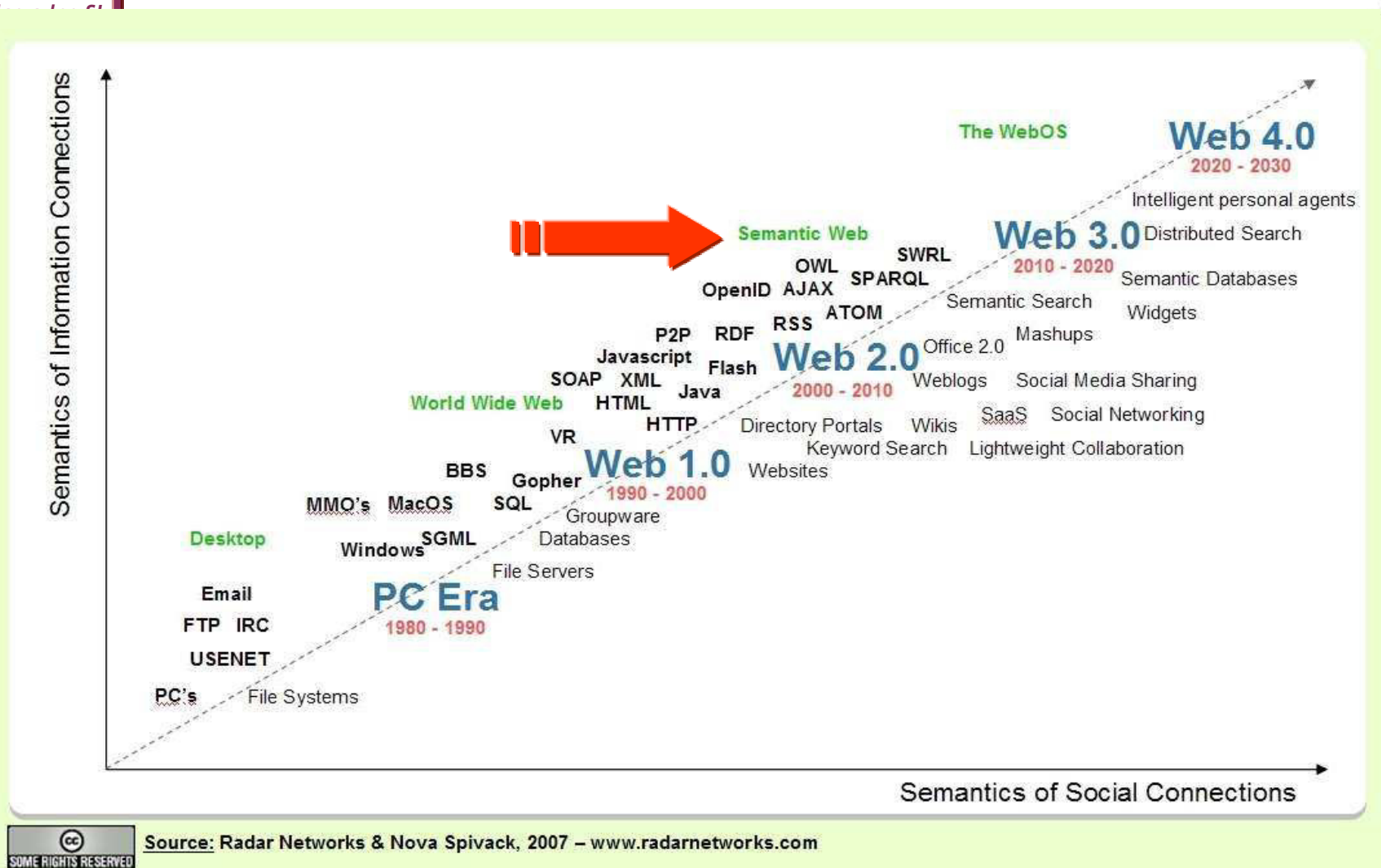
Bienvenido al 2.0

Technologies and trends shown in the word cloud:

- Mobility
- Diseño Web
- Usabilidad
- Sindicación
- Participación
- Ruby on Rails
- Estándares
- Ajax
- e-bussines
- Tagging
- Blogs
- Redes Sociales
- Comunidades
- Wikis
- Agregación de contenidos
- Podcast
- Web semántica
- WAI AA
- Accesibilidad

[VER MÁS +](#)

La Web del FUTURO (2007): Web x.0

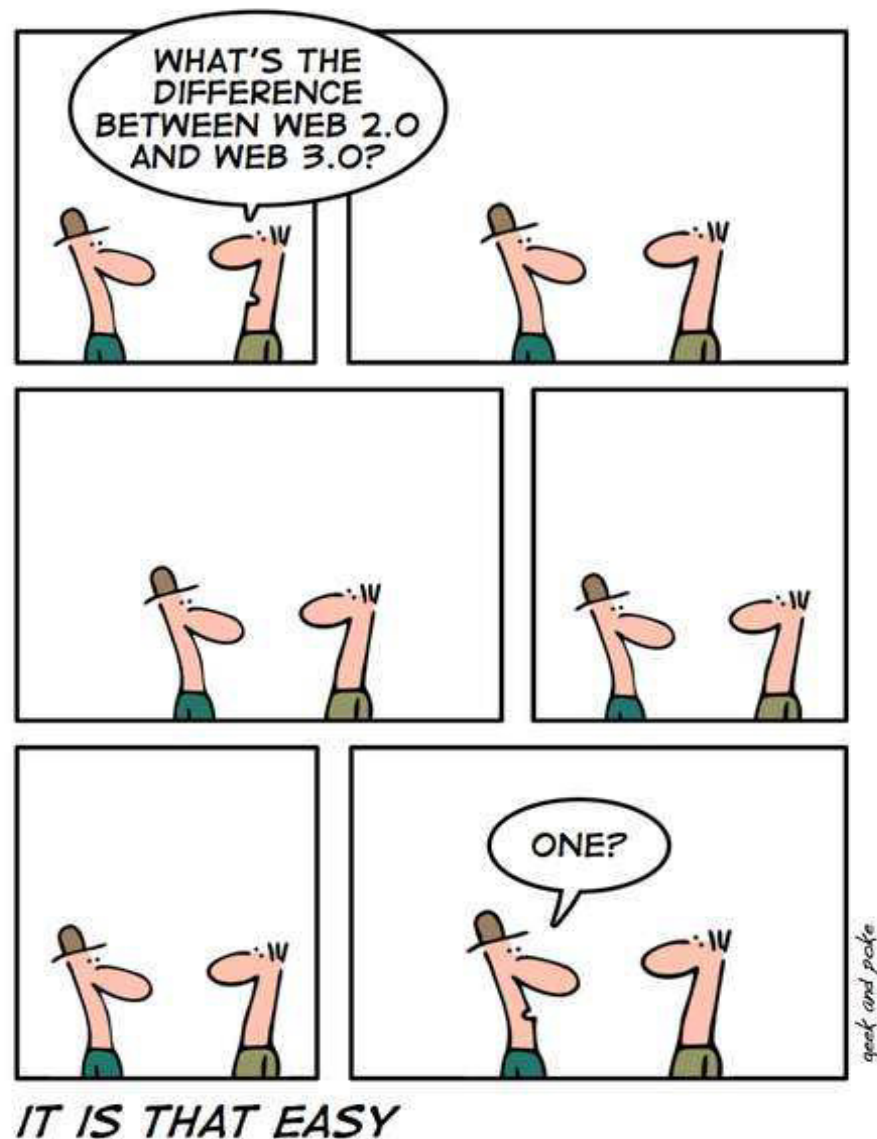


Fuente: a través de IRSweb (Javier Martínez Méndez):

<http://irsweb.blogspot.com/2007/03/cronologa-de-la-web.html>

La Web del futuro/presente: Web 3.0?

¡¡No es tan fácil!!





Web 2.0: Visión típica



Web 1.0 vs. 2.0: Comparación típica



DoubleClick
[Ofoto](#) (Kodakgallery), Altavista
[Akamai](#) (business online)
[mp3.com](#)
[Britannica Online](#)
Personal websites
Evite
Domain name speculation
Page views
Screen scraping
Publishing
CMS
Directories (*taxonomy*)
Stickiness

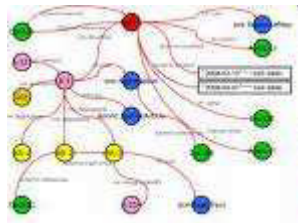
Google AdSense
[Flickr](#), [Corbis](#)
[BitTorrent](#) (cooperative distribution)
[Napster](#), [eMule](#), etc.
[Wikipedia](#)
Blogging
[Eventful](#) and [EVDB](#)
Search engine optimization
Cost per click
Web services
Participation
Wikis
Tagging ("*folksonomy*")
Syndication

Web 1.0

Web 2.0

Fuente: Adaptado de O'Reilly

Tim Berners-Lee, 2001: SW



“La Web Semántica (SW) no es una Web separada sino una extensión de la actual, en la cual la información se da con un significado bien definido, permitiendo que los ordenadores y las personas puedan trabajar juntos, en cooperación”.



Tim O'Reilly, 2006: Web 2.0



“El principio central que subyace a los gigantes nacidos en la Era de la Web 1.0, que han sobrevivido en la Era Web 2.0, parecen haber adoptado el poder de la Web para explotar la inteligencia colectiva”



Fuente: Traducción al español del artículo de O'Reilly: Qué es la Web 2.0:
<http://sociedaddelainformacion.telefonica.es/jsp/articulos/detalle.jsp?elem=2146>



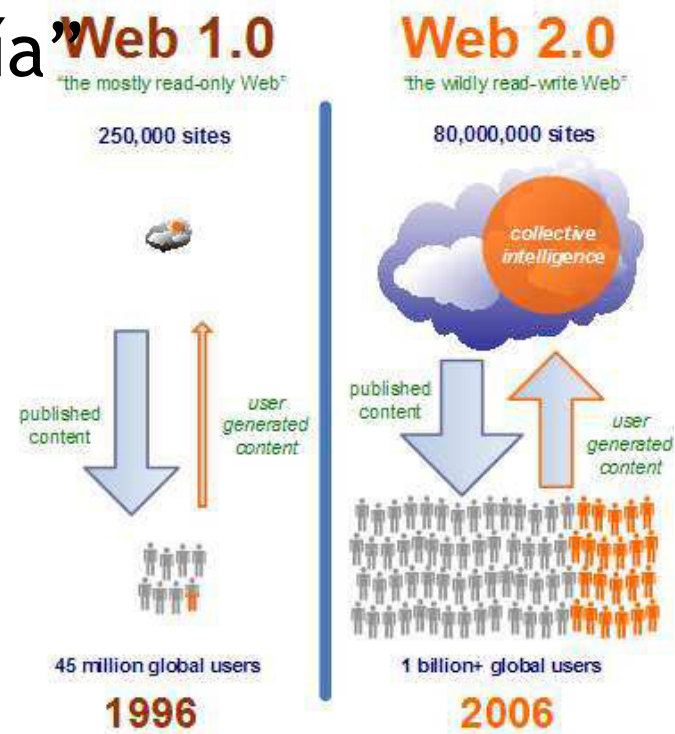
Web 2.0 es La Web Social (Hinchcliffe)



“La Web 2.0 tiene mucho más que ver con el cambio que se produce en las personas y en la sociedad que con la tecnología”

Algunos datos:

- Más de un billón de personas conectadas a Internet
- 100 millones de sitios Web
- Alrededor de 1/3 de adultos en USA han contribuido al contenido público de Internet (18% mayores de 65)



Fuente: <http://web2.wsj2.com/>

Tim Berners-Lee: noviembre 2006



“La Web no va de lo que podemos hacer con los ordenadores. Son personas, y sí, vale, están conectadas por computadores... Pero la informática, entendida como el estudio de lo que pasa en un ordenador, no te dice qué pasa en la Web.”



Fuente: New York Times
November 2, 2006



Pero... ¿qué es inteligencia colectiva?



- ¿Colecciones inteligentes?
 - *Bookmarking* (marcadores) colectivos, búsqueda colectiva, etc.
- “Base de datos de intenciones” (intenciones de búsqueda → actos voluntarios de *query*)
 - Verbos (acciones colectivas): *clicking, rating, tagging, buying*
- Lo que todos sabemos pero no nos atrevimos a decir en público antes:
 - Listas de discusión... y más Web 2.0: blogs, wikis, etc.



Tim Berners-Lee: Feb 2009 / Feb 2010



On the next Web (TED):

Datos enlazados

<http://www.w3.org/DesignIssues/LinkedData.html>

Nuevo TED de Tim Berners Lee en Febrero 2010:
No tiene desperdicio para analizar la evolución
de la Web y los datos enlazados:

http://www.ted.com/talks/tim_berniers_lee_the_year_open_data_went_worldwide.html

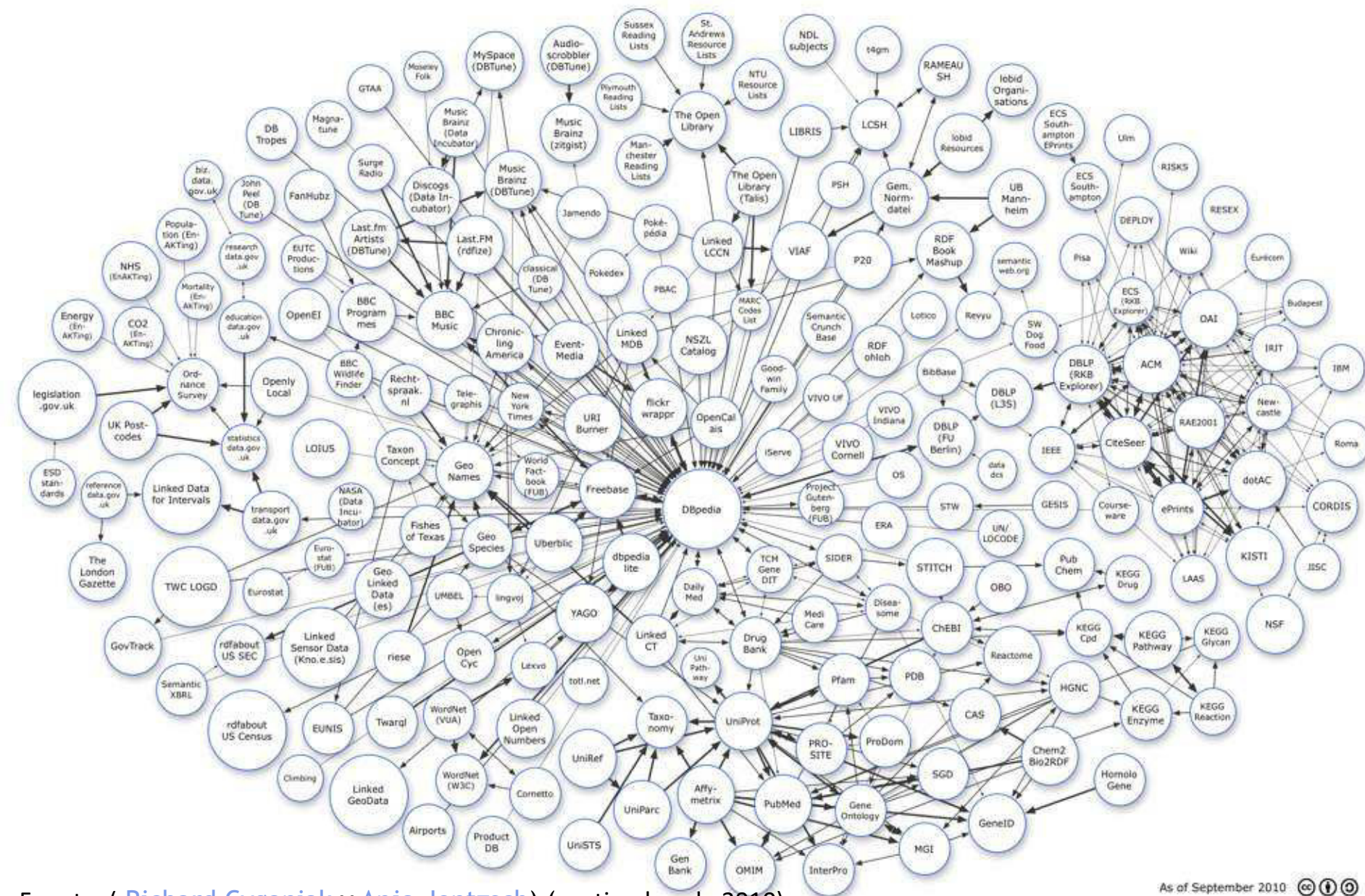


La Web del FUTURO (2010)



De islas semánticas a datos enlazados

Datos enlazados= *Linked Open Data*





De la Web Semántica a Datos Enlazados



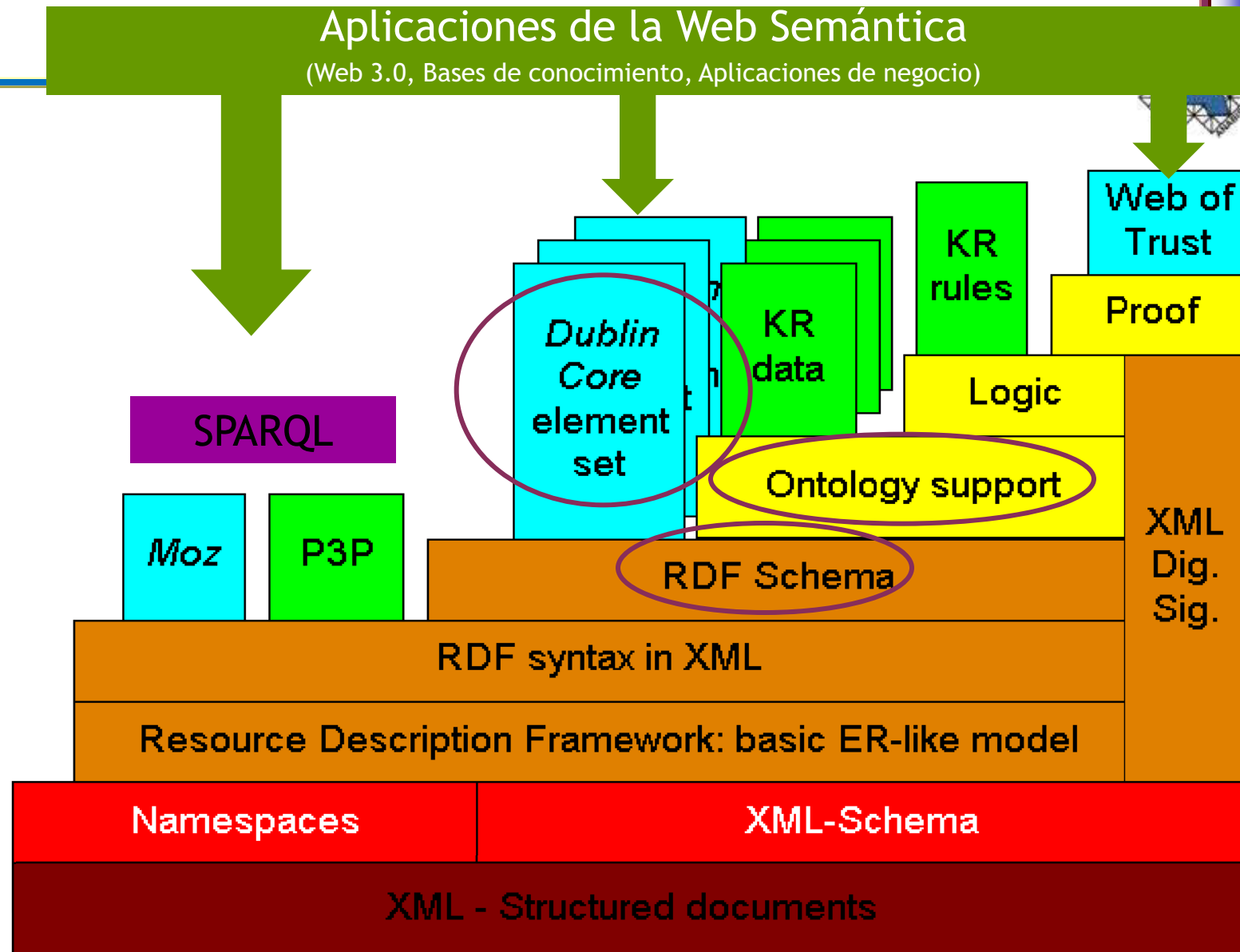
Coloquio Internacional
sobre Tecnologías aplicadas
a los Servicios de Información
"Hacia la Gestión Integral de Unidades de Información"
17-18 y 19/Noviembre, 2010

SW implica...



- **Lenguajes formales como:**
(DAML+OIL), RDF, OWL, etc.
- **Vocabularios:**
 - **Metadatos y esquemas de metadatos.**
 - **Ontologías:** tanto específicas de un dominio, como ontologías de alto nivel de tal forma que se proporcionen relaciones entre ontologías para la recuperación, el intercambio y la integración de la información interdisciplinar en la Web.





The Semantic Web Technology Stack (not a piece of cake...)

Most apps use only a subset of the stack

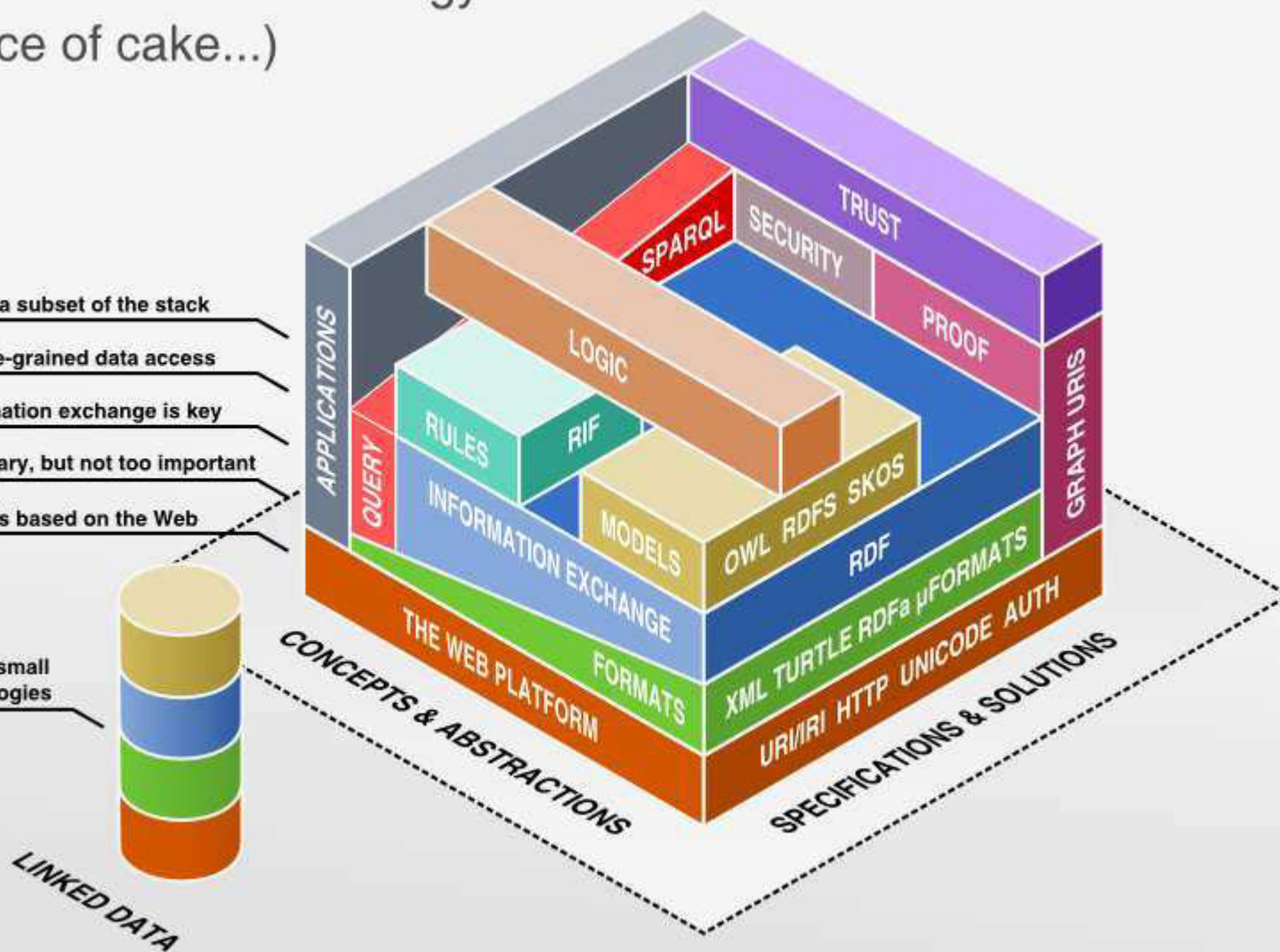
Querying allows fine-grained data access

Standardized information exchange is key

Formats are necessary, but not too important

The Semantic Web is based on the Web

Linked Data uses a small selection of technologies



El papel de los metadatos en Sistemas de información digital Sociales y Semánticos



Papel “tradicional”

- Identificación y descripción de objetos digitales
- Búsqueda y recuperación
- Ubicación de los documentos

Papel emergente

- Autoría y propiedad intelectual (DRM)
- Actualización de la información
- Preservación y conservación
- Limitación del uso
- Valoración del contenido (POWDER)
- Visibilidad de la información
- Formas de acceso
- Accesibilidad de los contenidos (WAI)
- Metadatos de usuario



Ontologías: representación de la información para la SW



- Son conjuntos de...
 - ... Conceptos o clases (términos que evocan conceptos)
 - ... Atributos (características de esos objetos)
 - ... **relaciones** entre los términos (conceptos),
 - ... que describen un dominio de aplicación concreto.
- Su objetivo es:
 - Creación de un diccionario de términos compartidos y comprensibles para diferentes aplicaciones y/o comunidades que permita la interoperabilidad.
 - Cualificar los términos para que representen conceptos
- Necesitan...
 - “Responsabilidad ontológica”
 - Formalización


Folksonomías: la indización social distribuida para la Web 2.0



- Def. **Folksonomía**, calco del inglés *folksonomy*, es un nuevo término que da nombre a la categorización colaborativa por medio de etiquetas simples en un espacio de nombres, sin jerarquías ni relaciones predeterminadas.
- Se trata de una práctica que se produce en entornos de Web social ([Web 2.0](#)) cuyos mejores exponentes son los sitios compartidos como [del.icio.us](#) (enlaces favoritos), [Flickr](#) (fotos), [Tagzania](#) (lugares), [flof](#) (lugares) o [43 Things](#) (proyectos, etc.).
- [Más sitios Web 2.0](#)



Folksonomías y social tagging

url	<input type="text" value="http://www.streampad.com"/>	
description	<input type="text" value="In tune with the world."/>	required 
notes	<input type="text" value="Streampad allows you to listen to your own music streamed to any computer!"/>	optional
tags	<input type="text" value="music streaming social"/>	space separated
suggestions	<input type="text" value="socialsoftware"/>	
<input type="button" value="save"/>		

▼ your tags (click to add)

» sort: alphabetically | by frequency

.net activity [ajax](#) [ambiance](#) [ambient](#) [answer](#) [appl3](#) [applet](#) [atesttag](#) [atom](#) [automation](#) [biology](#) [blog](#) [blu-ray](#) [books](#) [broadband](#)
[browser](#) [business](#) [cell](#) [cellphone](#) [charter](#) [chat](#) [chsarp](#) [clock](#) [computing](#) [coolashell](#) [copyright](#) [cryptography](#) [css](#) [culture](#) [danthesis](#)
[delicious](#) [design](#) [dvd](#) [encryption](#) [encyclopedia](#) [ergonomics](#) [fcc](#) [filesharing](#) [firefox](#) [fish](#) [flash](#) [flash8](#) [flickr](#) [foo+bar+](#) [future](#) [gadget](#)
[gadgets](#) [geocoder](#) [google](#) [government](#) [gps](#) [greasemonkey](#) [hd](#) [history](#) [humor](#) [hypocrisy](#) [im](#) [image](#) [images](#) [imaging](#) [immigration](#)
[information](#) [interopability](#) [ip](#) [ipod](#) [itp](#) [java](#) [javascript](#) [jjj](#) [knowledge](#) [kvm](#) [lawyer](#) [linksys](#) [linux](#) [mac](#) [map](#) [max](#) [mce](#) [media](#)
[message](#) [michigan](#) [mini](#) [mobile](#) [movie](#) [mozilla](#) [mp3](#) [mpx](#) [msp](#) [music](#) [MySQL](#) [navigation](#) [net](#) [networking](#) [nuclear](#) [onion](#)
[opensource](#) [oss](#) [p2p](#) [patent](#) [photo](#) [PHP](#) [physicalcomputing](#) [playlist](#) [pocketpc](#) [presence](#) [ps3](#) [psp](#) [quantum](#) [question](#) [radio](#)
[radiohead](#) [rap](#) [rdf](#) [reform](#) [rss](#) [satire](#) [seattleint](#) [security](#) [sharing](#) [SMS](#) [social](#) [socialsoftware](#) [software](#) [strategy](#) [sucks](#) [talk](#)
[technology](#) [text](#) [theory](#) [tools](#) [toy](#) [trademark](#) [tv](#) [visualization](#) [war](#) [weather](#) [web](#) [webdev](#) [webservice](#) [website](#) [windows](#) [wireless](#)
[work](#) [xbox](#) [xhtml](#) [xml](#) [xslt](#) [yahoo](#) 🎵

▼ network (click to add)

[for:dkan24](#) [for:jwhiting](#)



Steve museum: <http://steve.museum>



HELLO
MY NAME IS
Steve

STEVE: THE MUSEUM SOCIAL TAGGING PROJECT

Home

Research

Software

Steve in Action

FAQ

Links & Resources

Forum

Steve Tagger

search...

search

Welcome to the Steve Project

Steve is a collaboration of museum professionals and others who believe that social tagging may provide profound new ways to describe and access cultural heritage collections and encourage visitor engagement with collection objects. Our activities include researching social tagging and museum collections; developing open source software tools for tagging collections and managing tags; and engaging in discussion and outreach with members of the community who are interested in implementing social tagging for their own collections.


Steve as Research

The project team is engaged in systematic research into how social tagging can best serve the museum community and its visitors. Our current [research project](#), "T3: Text, Tags, Trust," a partnership with the University of Maryland's School, is funded, in part, by a National Leadership Grant for Research from the U.S. Institute of Museum and Library Services. [Results](#) of a recently-completed IMLS research project, "Researching Social Tagging and Folksonomy in the Art Museum," are presented in the Research section of this website. Data collected at <http://tagger.steve.museum> provides a testbed for researching our hypotheses about social tagging.

Steve as Software

To support our research, the steve team has developed a suite of open source tagging tools. The current version of the steve tagger tool is available at the project's [SourceForge](#) site, and members of the community are encouraged to download and install the tool. At its core, the tagger is a php web application that can publish a basic online collection and record tags. The project has issued a rich API that allows the integration of the tagger tool with existing tools and supports new development efforts. By itself, tagging is not technically difficult, but it is our intention to provide a easily-deployed common platform to allow users to share the costs of future tool development.

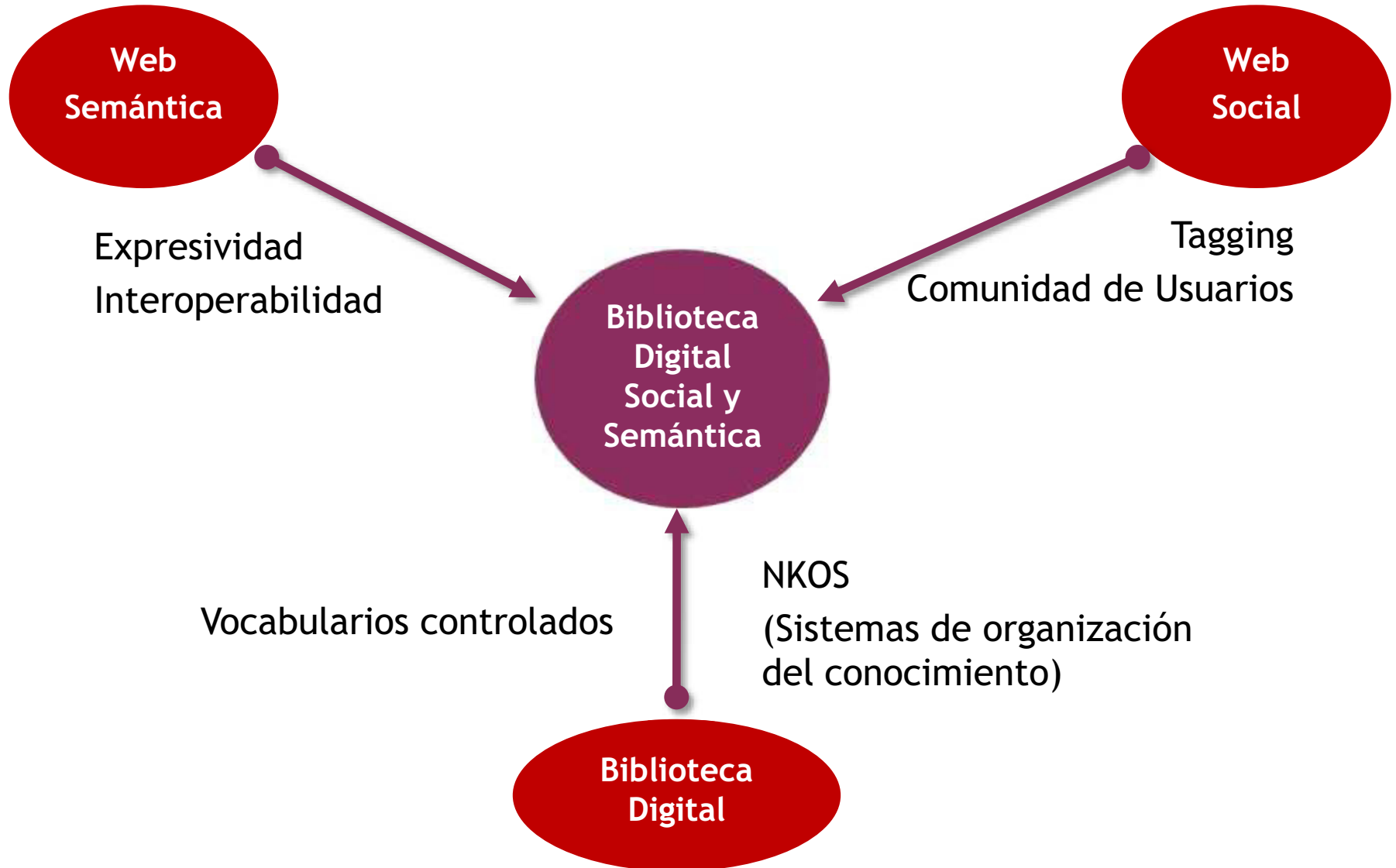
Steve in Action



wood
couple
primitive
seated
carving
affection



Sistemas de Información Digital Sociales y Semánticos: Nuevo paradigma



Algunos avisos del cambio de paradigma

Agenjo, Septiembre 2010



- Agenda de Lisboa 2010: Digital Libraries como iniciativa insignia ([i2010 - A European Information Society for growth and employment](#))
- Bibliotecas digitales como representación de una cultura ([Gallica](#))
- Respuesta europea a Google Books: [Europeana](#)
- Las grandes bibliotecas digitales: [Internet Archive](#), [Hathi Trust](#), [WDL](#), etc.
- Recolección de metadatos: los grandes agregadores: [OAlster](#), [OpenDOAR](#), etc.
- La unificación de todos los recursos digitales de un país ([DigitalNZ](#), [Australia TROVE](#)); [Catálogos NG](#)
- Las bibliotecas digitales tienen su sentido en la interrelación con otras, en su coordinación con otras.





A modo de conclusión: Reflexiones y recapitulación

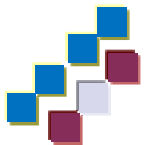


Web 1.0



La Web como “un portal de información”, la Web esaparate

- Exclusividad de la información
- Seccionar la WWW en directorios útiles
- Cada uno tenía su pequeño rincón en el ciberespacio
- Carece de:
 - Contexto
 - Interacción
 - Escalabilidad





La Web como plataforma

- Centrada en el poder de compartir (crear y validar)
- El poder de un sistema de organización más libre (tagging)
- Estableciendo conexiones para la integración futura (RSS)
- Carece de:
 - Personalización
 - Portabilidad verdadera
 - Interoperabilidad

Web 3.0 Meme Map

THE SEMANTIC WEB

Changing the web into a language that can be read and categorized by the system rather than humans



W3C

PERSONALIZATION

Contextualizing the web based on the people using it



+netvibes

ARTIFICIAL INTELLIGENCE

Extracting meaning from the way people interact with the web

Google
Suggest



MOBILITY

Everything, everywhere, all the time



- Web Semántica
- Web de datos
- Personalizable
- Ubicua
- Móvil

Esta es la Web en la que vamos a desarrollar nuestras Unidades de información de forma integral

¡¡¡Gracias por su atención!!!



¿Discusión? / ¿Preguntas?

emendez@bib.uc3m.es

San Cristobal 17 de Noviembre de 2010

